Supporting Information

Raskatov et al. 10.1073/pnas.1214267109

SI Text

RNA-Sequencing (RNA-Seq) Processing Needs to Take the Presence of Two Transcriptomes into Account. Initial attempts to map and quantify RNA-seq data using the commonly used software packages TopHat and Cufflinks (1) using just the human genome as a reference resulted in unsatisfactory results (which was demonstrated by control quantitative reverse transcription-PCR (qRT-PCR) experiments) likely due to the presence of significant amounts of host RNA in the libraries and the extensive sequence homology between highly conserved mouse and human genes. We investigated this by comparing the mapping efficiencies of the tumor samples to the human genome (hg19) to the mouse (mm9). The ratio between the mapping percentages to hg19 and mm9 is reflective of the tumor composition and was found to lay typically between 1.3 and 1.4, suggesting some 40% of the tumor mRNA to be mouse derived (Table S2). For comparison the ratio for the parent A549 cell line from an in vitro experiment was determined as 7.0, while that of a control mouse spleen sample was 0.12. This prompted us to devise the computational pipeline that simultaneously maps to both species' transcriptome as described in the main text.

Affected Genes Listed in Table 1 (Main Text). ***ATM: Official symbol: ATM and name: Ataxia telangiectasia mutated.*** The protein encoded by this gene belongs to the PI3/PI4-kinase family. This protein is an important cell cycle checkpoint kinase that phosphorylates; thus, it functions as a regulator of a wide variety of downstream proteins, including tumor suppressor proteins p53 and BRCA1, checkpoint kinase CHK2, checkpoint proteins RAD17 and RAD9, and DNA repair protein NBS1. This protein and the closely related kinase ataxia telangiectasia and Rad3 related are thought to be master controllers of cell cycle checkpoint signaling pathways that are required for cell response to DNA damage and for genome stability. Mutations in this gene are associated with ataxia telangiectasia, an autosomal recessive disorder. [provided by NCBI reference sequence (refSeq), Aug. 2010].

CCL2: Official symbol: CCL2 and name: Chemokine (C—C motif) ligand 2. This gene is one of several cytokine genes clustered on the q-arm of chromosome 17. Cytokines are a family of secreted proteins involved in immunoregulatory and inflammatory processes. The protein encoded by this gene is structurally related to the chemokine (C-X-C motif) subfamily of cytokines. Members of this subfamily are characterized by two cysteines separated by a single amino acid. This cytokine displays chemotactic activity for monocytes and basophils but not for neutrophils or eosinophils. It has been implicated in the pathogenesis of diseases characterized by monocytic infiltrates, like psoriasis, rheumatoid arthritis, and atherosclerosis. It binds to chemokine receptors CCR2 and CCR4. (provided by refSeq, July 2008).

EGFR: Official symbol: EGFR and name: Epidermal growth factor receptor. The protein encoded by this gene is a transmembrane glycoprotein that is a member of the protein kinase superfamily. This protein is a receptor for members of the epidermal growth factor family. EGFR is a cell surface protein that binds to epidermal growth factor. Binding of the protein to a ligand induces receptor dimerization and tyrosine autophosphorylation and leads to cell proliferation. Mutations in this gene are associated with lung cancer. Multiple alternatively spliced transcript variants that encode different protein isoforms have been found for this gene. (provided by refSeq, July 2010).

MMP28: Official symbol: MMP28 and name: Matrix metallopeptidase 28. Proteins of the matrix metalloproteinase family are involved in the breakdown of extracellular matrix for both normal physiological processes, such as embryonic development, reproduction and tissue remodeling, and disease processes, such as asthma and metastasis. This gene encodes a secreted enzyme that degrades casein. Its expression pattern suggests that it plays a role in tissue homeostasis and in wound repair. Transcript variants encoding different isoforms have been described. (provided by refSeq, March 2010).

NPTX1: Official symbol: NPTX1 and name: Neuronal _Pentraxin I. NPTX1 is a member of the neuronal pentraxin gene family. Neuronal pentraxin 1 is similar to the rat NP1 gene that encodes a binding protein for the snake venom toxin taipoxin. Human NPTX1 mRNA is exclusively localized to the nervous system. (provided by refSeq, July 2008).

ROBO1: Official symbol: ROBO1 and name: Roundabout, axon guidance receptor, homolog 1 (drosophila). Bilateral symmetric nervous systems have special midline structures that establish a partition between the two mirror image halves. Some axons project toward and across the midline in response to long-range chemoattractants emanating from the midline. The product of this gene is a member of the immunoglobulin gene superfamily and encodes an integral membrane protein that functions in axon guidance and neuronal precursor cell migration. This receptor is activated by *SLIT*-family proteins, resulting in a repulsive effect on glioma cell guidance in the developing brain. A related gene is located at an adjacent region on chromosome 3. Multiple transcript variants encoding different isoforms have been found for this gene. (provided by refSeq, March 2009).

SERPINE1: Official symbol: SERPINE1 and name: Serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1. This gene encodes a member of the serine proteinase inhibitor (serpin) superfamily. This member is the principal inhibitor of tissue plasminogen activator and urokinase, and hence is an inhibitor of fibrinolysis. Defects in this gene are the cause of plasminogen activator inhibitor-1 deficiency (PAI-1 deficiency), and high concentrations of the gene product are associated with thrombophilia. Alternatively spliced transcript variants encoding different isoforms have been found for this gene. (provided by refSeq, Sep. 2009).

1. Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515.

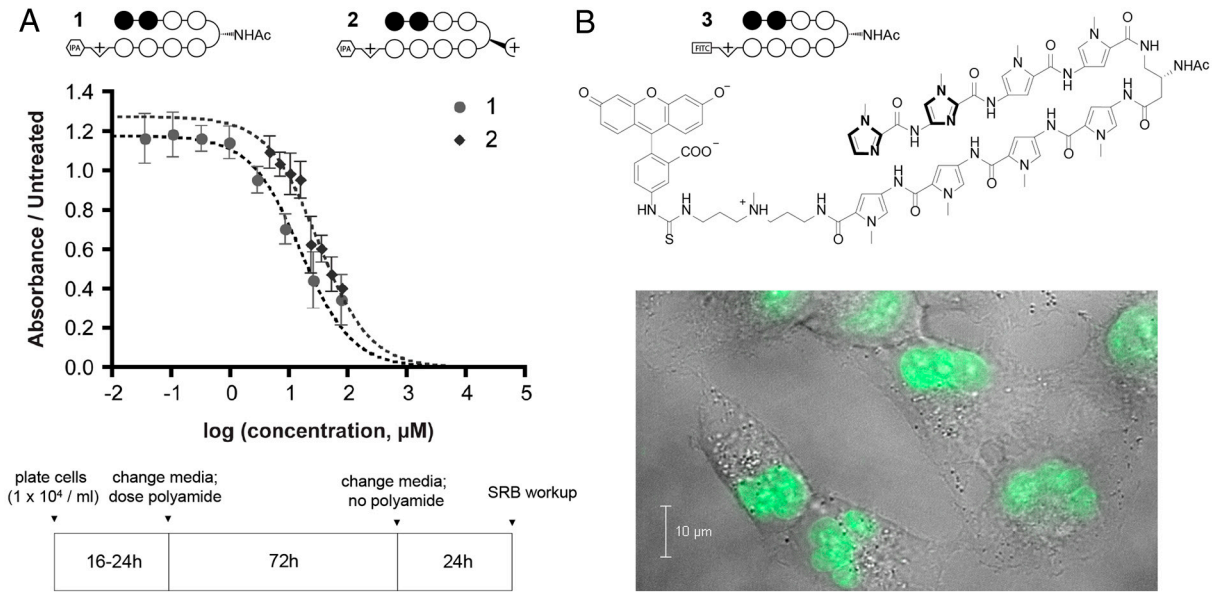
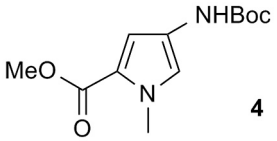


Fig. S1. (*A*) In vitro cytotoxicity of **1** *vs* **2** against the A549-luc-C8 cell line, measured employing the sulforhodamine B (SRB) assay. For full chemical structures of **1** and **2** see Fig. 1 of the main text. (*B*) Uptake of **3** in A549-luc-C8, measured by confocal microscopy (LSM 5 Exciter). Cells were incubated with Py-Im polyamide **3** at 1 uM final concentration for 12 h, washed once with PBS and imaged.



4

Fig. S2. Internal standard **4**, used for calibration of analytical HPLC.

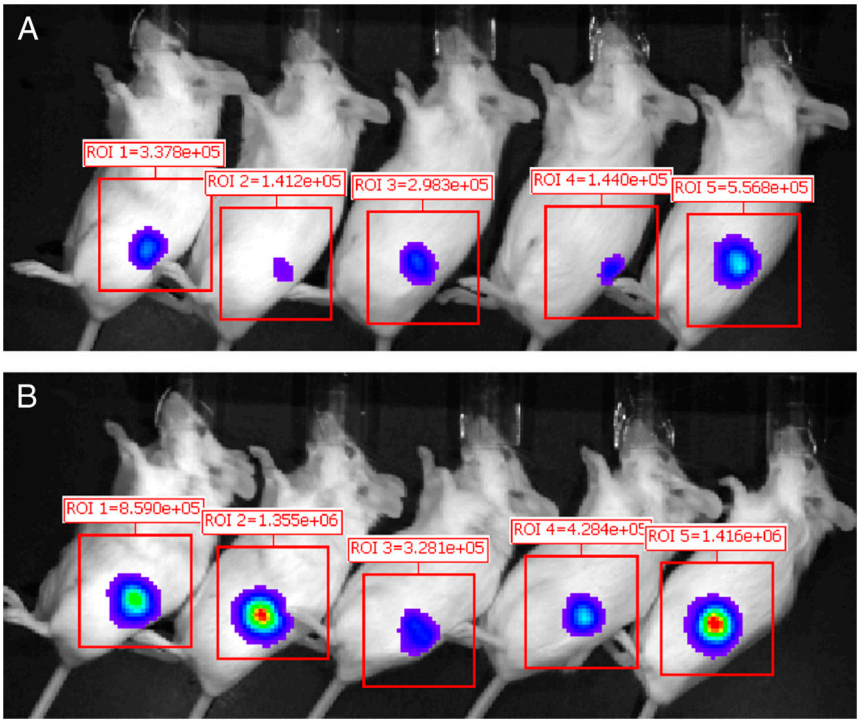


Fig. S3. (*A*) Vehicle group. (*B*) Treated with **1** as outlined in the experimental section. Averaged luciferase output values with the associated standard deviations are $(3.0 \pm 1.7) \times 10^5$ for vehicle and $(8.8 \pm 5.1) \times 10^5$ for treated.

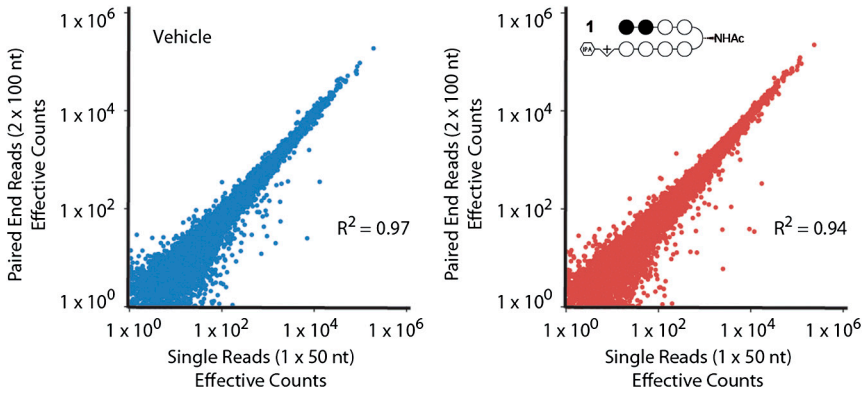


Fig. S4. Correlation between single reads (1 × 50 nt) and paired-end reads (2 × 100 nt) sequencing for tumors obtained from vehicle- (*Left*) and Py-Im polyamide-treated (*Right*) animals.

Table S1. Treatment schedule for SCID-beige animals with Py-Im polyamide 1. A549-luc-C8 were engrafted at 5×10^6 cells in 200 μL media per animal. Compound 1 was injected at 120 nmol/animal in 200 μL vehicle (PBS/DMSO, 4:1) per injection

Day	0	1	2	3	4	5	6	7	8	9	10
Engraft	X										
Treat	X		X		X			X		X	
Weigh	X						X				
Image											X
Harvest											X

Table S2. Mapping percentages of the A549-luc-C8 xenograft-derived RNA-seq reads to hg19 and mm9 (entries 1–6). Mapping percentages of the parent A549 cell line (entry 7) and SCID-bg mouse spleen mRNA (entry 8) are given for reference. All RNA-seq experiments were conducted with 1 × 50 nt (single reads)

Entry	Condition	No. of reads	% hg19	% mm9	[% hg19]/[% mm9]
1	XenoVehicle 1	35478968	59.5	42.9	**1.39**
2	Xeno Vehicle 2	50839514	36.5	27.5	**1.33**
3	Xeno Vehicle 3	50150429	59.2	46.3	**1.28**
4	Xeno Treated 1	54437744	59.5	43.3	**1.37**
5	Xeno Treated 2	49087273	39.7	28.9	**1.38**
6	Xeno Treated 3	34553534	60.4	44.4	**1.36**
7	A549 in vitro	35187689	83.0	11.9	**6.97**
8	SCID-bg spleen	34932537	11.9	95.8	**0.12**

Table S3. qRT-PCR measurements of expression changes of genes shown in Fig. 5 (A549-luc-C8 cell line); brackets indicate gene upregulation upon treatment. Xenograft: three independent experiments with $N = 5$ animals per treatment condition (vehicle vs 1) were averaged. Cell culture: where indicated, the cells were incubated with Py-Im polyamide 1 at 10 μM final concentration in 0.1% DMSO as vehicle

Gene	Xenograft	Culture, 48 h	Culture, 72 h
ATM	1.5 ± 0.2	1.9 ± 0.2	2.5 ± 0.3
NPTX1	3.3 ± 0.6	2.6 ± 0.5	3.5 ± 1.15
ROBO1	1.5 ± 0.2	3.3 ± 0.5	4.5 ± 0.7
MMP28	[1.5 ± 0.3]	1.1 ± 0.75	1.1 ± 0.7
EGFR	1.2 ± 0.2	1.9 ± 0.2	2.8 ± 0.3
CCL2	2.3 ± 0.4	2.2 ± 0.3	4.4 ± 0.45
SERPINE1	2.0 ± 0.2	8.3 ± 1.0	15.7 ± 1.65